



# CLUSTERING ALGORITHM USED FOR CYBER CRIME DATA

Nagendra Kumar Sahu<sup>1</sup> | Satish Kumar Sahu<sup>1</sup>

<sup>1</sup> Research Scholar, Mats University Raipur (C.G.).

## ABSTRACT

Cybercrime is becoming ever more serious. Findings from the 2002 Computer Crime and Security Survey show an upward trend that demonstrates a need for a timely review of existing approaches to fighting this new phenomenon in the information age. In this paper, we define different types of cybercrime and review previous research and current status of fighting cybercrime in different countries that rely on legal, organizational, and technological approaches. We focus on a case study of fighting cybercrime in India and discuss problems faced. Finally, we propose several recommendations to advance the work of fighting cybercrime.

**KEYWORDS:** Reason of Cyber Crime, Methodology, K-min algorithm, Result analysis.

## INTRODUCTION:

Cybercrime is criminal activity done using computers and the Internet. This includes anything from downloading illegal music files to stealing millions of dollars from online bank accounts. Cybercrime also includes non-monetary offenses, such as creating and distributing viruses on other computers or posting confidential business information on the Internet.

### Reasons for Cybercrime:

1. *Capacity to store data in comparatively small space:* The computer has unique characteristic of storing data in a very small space. This affords to remove or derive information either through physical or virtual medium makes it much easier.
2. *Easy to access:* The problem encountered in guarding a computer system from unauthorized access is that there is every possibility of breach not due to human error but due to the complex technology. By secretly implanted logic bomb, key loggers that can steal access codes, advanced voice recorders; retina imagers etc. that can fool biometric systems and bypass firewalls can be utilized to get past many a security system.
3. *Complex:* The computers work on operating systems and these operating systems in turn are composed of millions of codes. Human mind is fallible and it is not possible that there might not be a lapse at any stage. The cyber criminals take advantage of these lacunas and penetrate into the computer system.
4. *Negligence:* Negligence is very closely connected with human conduct. It is therefore very probable that while protecting the computer system there might be any negligence, which in turn provides a cyber criminal to gain access and control over the computer system.
5. *Loss of evidence:* Loss of evidence is a very common & obvious problem as all the data are routinely destroyed. Further collection of data outside the territorial extent also paralyses this system of crime investigation.
6. Cybercriminals have embedded malware into legitimate applications and they're targeting poorly secured Wi-Fi spots, stealing passwords, and more in their quest to steal information.
7. Attackers like to exploit unauthorized products with weak security controls in the corporate cloud.
8. Zombie servers that are left unattended or are not updated offer additional ways to access networks.
9. Known vulnerabilities are not patched in time – a study has found that it takes an average of 193 days before a patch gets applied to fix a problem, even though the patch is available. That's an easy place for exploiters to get in and do their damage.

## RELATED WORK:

It is important to read previous related work to both learn from the experience of others and to add something to our existing body of knowledge. Existing literature has been reviewed in 3 different areas:

### Crime data mining, data extraction and data focus:

**Cyber Crime Data Mining:** Data mining is defined as the discovery of interesting Structure in data, where structure designates patterns, statistical or predictive

models of the data, and relationships among parts of the data. The Data mining techniques is using for some results on crime mining. This Technique is applied to study Cyber Crime cases, which mainly concerned entity extraction, pattern clustering, classification and social network analysis.

**Data Extraction Event:** In web page extraction is the process to extract attributes and relationship. The Idea of this event extraction is the method of retrieving the information from database. The Proposed a method to append events for the concept of data mining techniques. Entity Extraction has been used to automatically identify person, address,

**Data Focus:** This research paper is on web mining of content, using clustering techniques the web mining focus on the text. In data focus clustering will convert nonlinear statistical relationship between high dimensional data into simple geometrical relationship in low dimensional display

## METHODOLOGY:

**Data Collection:** The data set is the collection of field in the data from web pages on the internet. The data set which consists of the text from web pages and the pictures, videos or sound format will be ignored.

**Preprocessing:** A data preprocessing is a process that consists of data cleaning, data integration and data transformation which is usually processed by a computer program. It intends to reduce some noises, incomplete and inconsistent data. The Results from preprocessing step can be later proceeding by data mining algorithm. Clustering techniques are used to fetch the information of criminals

**Data Mining:** Data mining deals with the discovery of unexpected patterns and new rules that are "hidden" in large databases. The use of data mining in this paper is to give the structured data from unstructured data of judge. In this paper the Data Mining techniques of crime in two directions they are:

1. Classification of Cyber Crime
2. Clustering Technique of Cyber Crime

1. **Classification of Cyber Crime:** Crime is defined as "an act or the commission of an act that is forbidden, or the omission of a duty that is commanded by a public law and that makes the offender liable to punishment by that law". Crime is referred to as a comprehensive concept that is defined in both legal and non-legal sense.

2. **Clustering Technique of Cyber Crime:** Data clustering is a process of putting similar data into groups. A clustering algorithm partitions a data set into several groups such that the similarity within a group is larger than among groups. Clustering Can also be considered the most important unsupervised learning technique; so, as every other problem of this kind, it deals with finding a structure in a collection of unlabeled data. There are so many techniques used in clustering,

**K-Means Clustering Algorithm:** K-means algorithm mainly used to partition the clusters based on their means. Initially number of objects are grouped and specified as K clusters. The algorithm clusters observations into K groups, where K is provided as an input parameter. It then assigns each observation to clusters based upon the observation proximity to the mean of the cluster. The cluster's mean is then recomputed and the process begins again. In this paper the use of K means algorithm is the process of getting a structured data from an unstructured data. The Working of algorithm is

explained as follows:

k : pre-determined number of clusters

Algorithm (Step 0: determine value of k)

**Step 1:** Randomly generate k random points as initial cluster centers

**Step 2:** Assign each point to the nearest cluster center

**Step 3:** Re-compute the new cluster centers

**Repetition step:** Repeat steps 3 and 4 until some convergence criterion is met (Usually that the assignment of points to clusters becomes stable).

#### RESULT:

The result of this crime management system is given below: Judgment of a crime

**Table 1: Unstructured data**

Case no 0727 ABC is sentenced to jail for 15 years in the Raipur for the Cheating crime of a person using Computer

**Table 2: Structured data**

Case no	0727
Name	ABC
Crime Type	Cheating
Act	66D
Judgment	15 years
Location	Raipur

The use of K-means clustering algorithm in this paper is defining the structured data from unstructured data from data base. The prediction of this retrieving information from database is clear and understands to the user of the system. Predict the name, case, judgment, etc. from the judge. The Judge gives the unstructured information of a criminal as judgment. The prisoner admin get the data of particular criminal information in a structured format.

#### CONCLUSION:

Cyber Crime data is a sensitive domain where effective data mining techniques play a vital role for Cyber Crime analysis. In this paper the classify and cluster techniques are used to analyses the crime data from database. This technique is faster to get the data through web; the effective web mining is to get the unstructured data to structured data. The classifications of crime type are Cheating, Forgery, etc. and clustering of crime using k-means to retrieve the data.

#### REFERENCES:

1. R.G Uthra Emerging Trends in Utilization of Data Mining in Criminal Investigation: An Overview, Journal of Data Mining Technique to Analyze Crime Data, International Journal for Technological Research in Engineering , 2013
2. Malathi. A, Dr. S. Santhosh Baboo, Anbarasi. A, An Intelligent Analysis of a City Crime Data Using Data Mining, Internation Conference on Information and Electronics Engineering, 2011
3. Devesh Bajpai, Emerging Trends in Utilization of Data Mining in Criminal Investigation: An Overview, Journal Of Environmental Science, Computer Science And Engineering & Technology, 2012
4. Anshu Sharma, Shilpa Sharma, An Intelligent Analysis of web Crime Data Using Data Mining, Internauton Journal of Engineering and Innovative Technology, 2012.
5. [www.newworldencyclopedia.org/entry/Cybercrime](http://www.newworldencyclopedia.org/entry/Cybercrime)
6. [www.bukisa.com/articles/206\\_internet-security-concepts](http://www.bukisa.com/articles/206_internet-security-concepts)
7. <http://en.wikipedia.org/wiki/Computercrime>
8. <http://www.wisegeek.com/what-is-cybercrime.htm>
9. How to Prevent Cyber Crime | eHow.com  
[http://www.ehow.com/how\\_4967690\\_prevent-cyber-Crime.html](http://www.ehow.com/how_4967690_prevent-cyber-Crime.html)